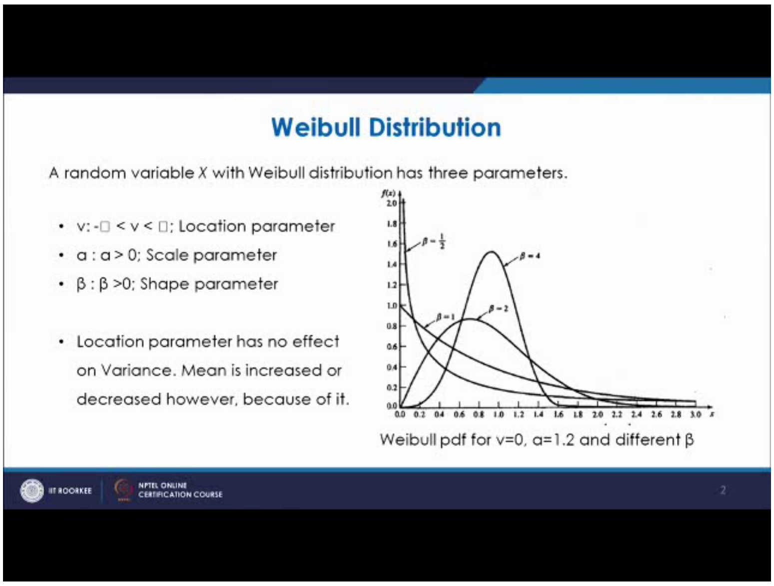
Modeling & Simulation of Discrete Event Systems
Dr. Pradeep K Jha
Department of Mechanical and Industrial Engineering
Indian Institute of Technology, Roorkee

Lecture – 09
Continuous Distribution Functions and Empirical Distribution Functions

Welcome to the lecture on Continuous Distribution Functions and Empirical Distribution Functions. So, in the last lecture we discussed about few of the continuous distribution functions and further we will discuss about some of the other important continuous distribution functions and also we will discuss about the empirical distribution functions. So, in the least what we discussed in the last class ahead of that the first distribution function which will discussed is the Weibull distribution function.

(Refer Slide Time: 00:54)

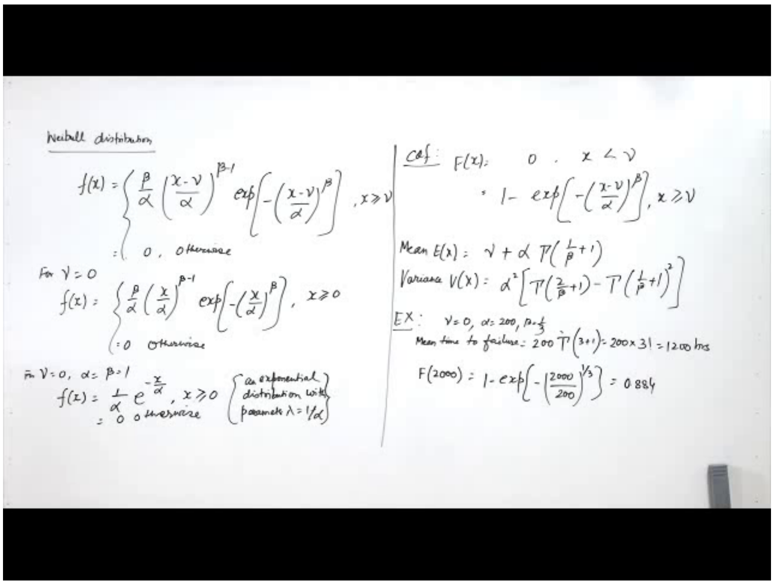


So, in the Weibull distribution function the random variable x which has which follows the Weibull distribution it has 3 parameters one is location parameter, one is the scale parameter and another is the shape parameter.

So, as we see the Weibull distribution function curves is shown like this and you have the 3 parameters nu this is minus infinity to plus infinity this is infinity here. So, new varies from minus infinity to plus infinity that is location parameter it will tell about the place at which the function will start, if the nu is 0 this function will start here at 0. So, it will be the starting point. So, that is your nu the significance of nu. Alpha is the scale parameter and beta is the shape parameter. So, depending upon the different value of alpha and beta you have the different shapes of these Weibull distribution.

Now, in this case the probability distribution function takes certain form and the f x, for Weibull distribution.

(Refer Slide Time: 02:26)



For Weibull distribution f x is beta by alpha x minus nu by alpha raised to the power beta minus 1 then exponential minus of x minus nu by alpha raised to the power beta. So, the value of the probability for any value of x will be computed using this formula where x has to be more than nu. So, x has to be more than equal to nu. So, when x is more than equal to nu as we see here the x value has to be more than or equal to nu and at that point the f x value will be computed using that and if it is not so if it is less than that then it will be 0 otherwise. So, this is how the f x is computed.

Now, in that what we see is if you compute this if we find the different forms of this expression and if you take nu as 0, so for location parameter value as 0 what we see is f x will be you can see here. So, we have to put like this, beta by alpha and then x by alpha because nu becomes 0. So, it will be beta minus 1 then exponential minus of x by alpha raised to the power beta. So, that will be x more than equal to 0 and then it will be 0 otherwise. If it is less than equal to 0 in that case it will take this. For further if we take nu equal to 0 and alpha and beta as 1 in that case the function comes like this f x you can write it as 1 by alpha if we express in terms of alpha. So, it will be 1 by alpha exponential minus x by alpha. So, that is what we see and this will be for x more than equal to 0 and 0 otherwise. So, it is nothing, but an exponential distribution with parameter 1 by alpha.

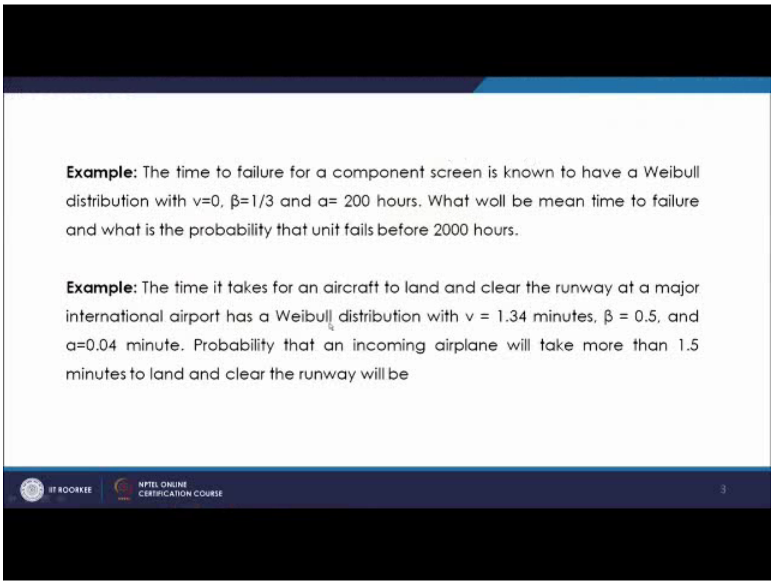
So, once we have that now what we see from this curve you can see if you have different values of nu, alpha and beta. So, for different value of beta basically nu and alpha is fixed in that case you have this kind of curve which is generated and this is known as Weibull distribution.

Further if we try to compute the cumulative distribution function value for this. So, cdf for this comes like F x will be f x will be 0 if x is less than nu. So, as we know from the location parameter before that the cumulative probability value will be 0. So, cumulative distribution function value becomes 0. And then for the value which is more than nu it will be one minus exponential minus of x minus nu by alpha raise to the power beta.

So, this will be x more than equal to nu. So, this will be x more than equal to nu. So, this is how you can find the cumulative probability when any event is said to be Weibull distributed. Now for this Weibull distribution the mean and variance are computed. So, mean E x this is computed to be equal to nu plus alpha gamma of 1 by beta plus 1. So, this gamma function we have already discussed gamma of beta will be beta minus 1 factorial. So, that is how we will compute this gamma value and the variance v x it is coming out to be alpha square into gamma of 2 by beta plus 1 minus gamma of 1 by beta plus 1 square. So, this is how the mean and variance is also calculated in the case of such Weibull distribution.

What we see is we see here that the variance expression does not have nu. So, the variance does not depend upon nu, so that is also clear from this graph the variance or a spread does not have any effect I mean it does not get effected because of any value of nu, but mean or mode. So, as we see the mean is increased or decreased because of that as we see if the nu is different the mean will be different because of the placement of these curves to different you know on that axis, this mean will we certainly changing. So, let us see some of the examples.

(Refer Slide Time: 09:26)



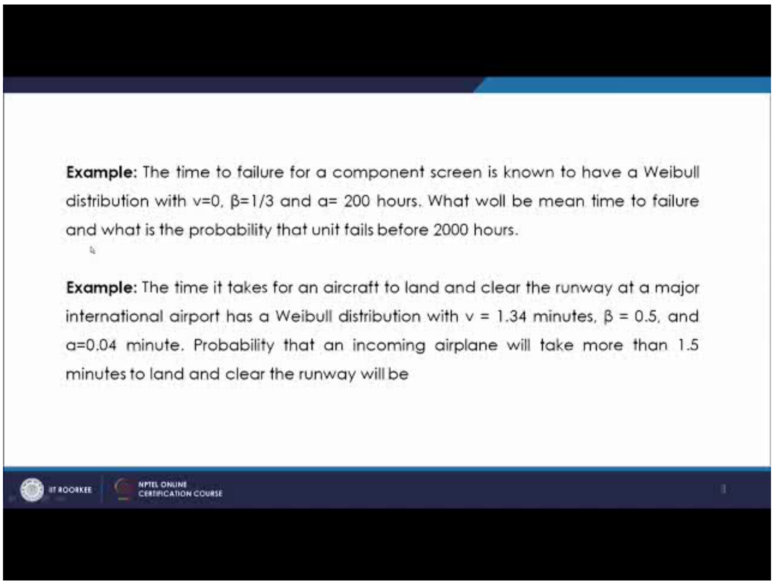
So, like one of the example is that the time to failure for a components screen is known to have a Weibull distribution with nu equal to 0 beta 1 by 2 and alpha is 200 hours. So, for such problem it is written that the screen is said to be Weibull distribution. So, the time to failure is having Weibull distribution and the 3 parameters are given. So, what will be the mean time to failure? So, mean time to failure if you try to find. If you try to find the mean time to failure and the parameter values are given nu equal to 0 and alpha is 200 and beta is 1 by 3.

So, we can get from here directly mean time to failure it will be nu plus, nu 0. So, it will be then alpha times. So, alpha is 200 times gamma of 1 by beta. So, 1 by beta is 3 plus 1 gamma 4 and gamma 4 is 3 factorial 3 factorial is 6, so it will be 1200 hours. So, mean time to failure for such a distribution with these parameter values will be 1200 hours that is how we calculate it. It is again written that what is the probability that the unit fails before 2000 hours. So, it has you have to find that probability so that it fails before 2000 hours that is cumulative probability value you have to find and for that you have to find the F 2000 this value will tell the probability that it will fail before 2000 hours and that will be again 1 minus exponential x will be 2000. So, 2000 upon alpha is 200 and raised to the power beta. So, 1 by 3 and if you get this value you will be getting 0.884. So, this is how you can calculate these probability values for such distribution functions.

Then next the probability distribution function which will come will be the triangular distribution function.

Now as we see in the case of triangular distribution functions you have 3 parameters a b and c.

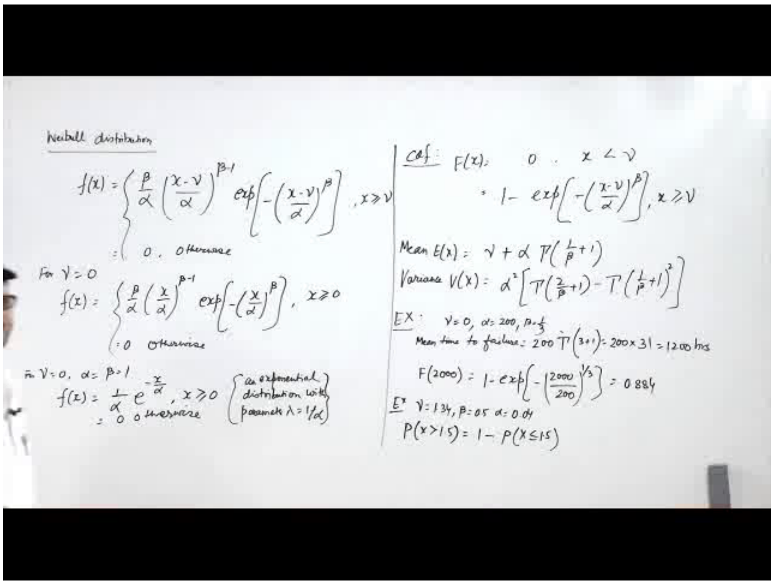
(Refer Slide Time: 12:22)



So, before that we have also to see we have another question the another question which is that the time it takes for an aircraft to land and clear the runway at a major international airport has a Weibull distribution with nu equal to 1.34 minutes, beta 0.5 and alpha is 0.04 minutes. So, as we see that the time for landing and clearing that happens on the you know airport that you have the time in which it will come it will land and then it will clear this time is basically said to be Weibull distributed with nu as 1.34 minutes. So, it will start from 1.34 minutes beta and alpha value is given. Now for that probability is to be found out where the incoming plane will take more than 1.5 minutes to land and clear the runway.

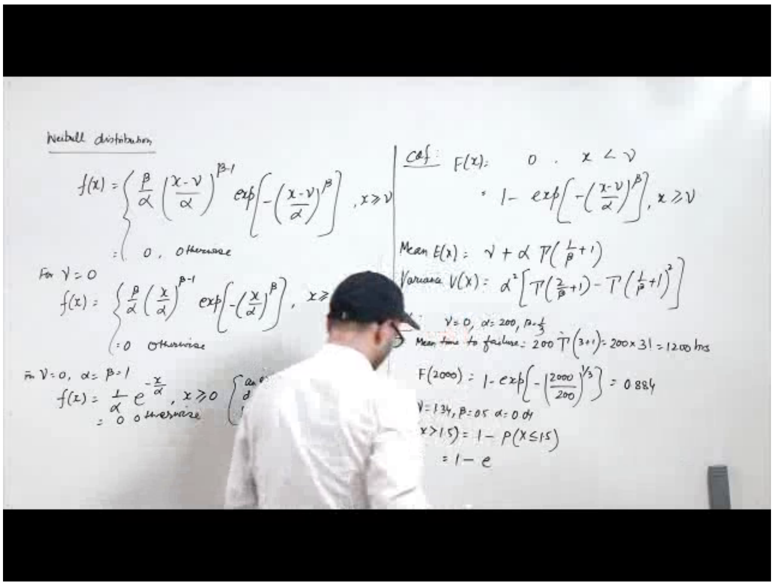
So, in this case you want to find the probability when the plane will take more than 1.5 minutes. So, in that case as we see we have nu as 1.34. So, further example is nu as 1.34 beta and alpha is given beta is given as 0.5 and alpha is 0.04.

(Refer Slide Time: 13:45)



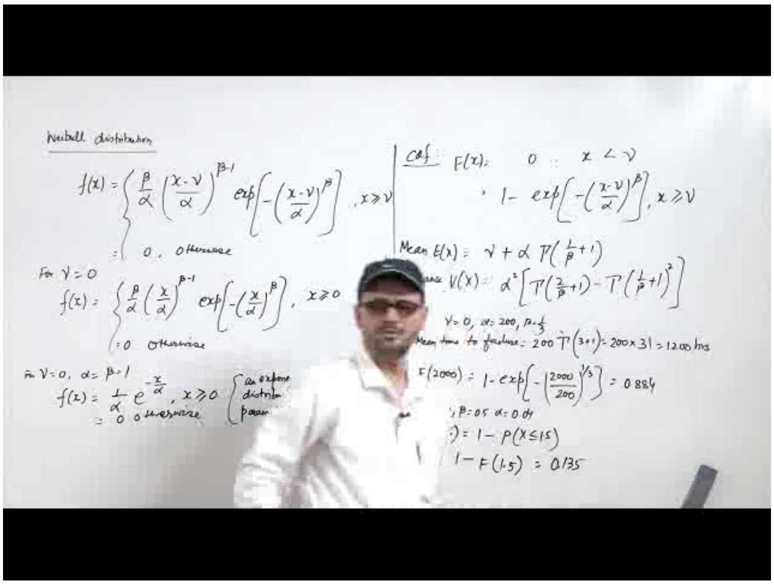
So, probability you have to find when it will be you know taking more than 1.5 minutes. So, it will be x is more than 1.5 that will be 1 minus p x will be less than equal to 1.5. So, we are subtracting it from one that probability where it will take less than 1.5 times and for that this is nothing, but the cumulative probability value. So, it will you will get from cdf that is f x. So, it will be 1 minus f x as we know, so f of 1.5.

(Refer Slide Time: 14:37)



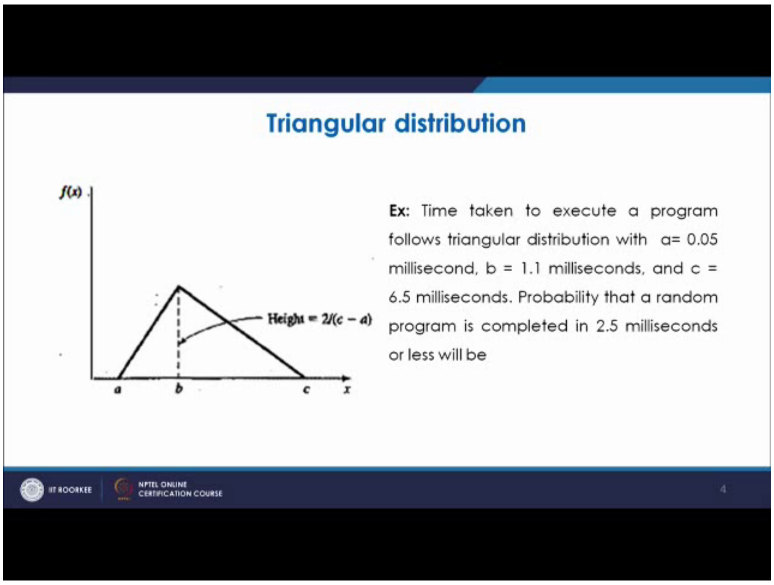
So, this value if you calculate and finally, the answer comes out to be 0.135.

(Refer Slide Time: 14:56)



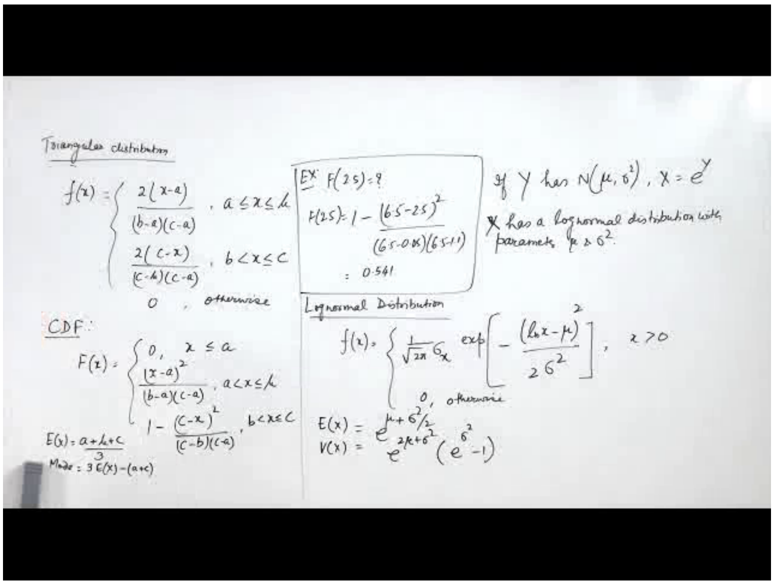
Next we have the triangular distribution.

(Refer Slide Time: 15:02)



So, as we see in the case of triangular distribution we have 3 points that is a b and c. So, this a b and c has the effect on the probability values and for that the distribution function looks like this.

(Refer Slide Time: 15:28)



So, for a triangular distribution f x will be 2 into x minus a upon b minus a c minus a. So, as we see in this case you have 2 zones one is from a to b and another is from b to c as it is shown that it is height is 2 upon c minus a. So, you will have the probability of happening between a and b or between b and c. So, when the x value lies between this region a and b in that case the probability value or f x value will be computed using this formula. So, here a it is in between a and b. So, when it goes from b to c that is in the second reason in this reason the f x value will be 2 into c minus x upon c minus b into c minus a. So, for that it has to be more than b and less than equal to c. So, it is b to less than equal to c and then if it is beyond that it will be 0, so that will be 0 otherwise. So, this is how the pdf looks like for this triangular distributions.

For them the cumulative distribution function is taking in this form f x will be 0 if x is less than equal to a because before a it does not have any probability values so it is 0. Then if it is between a and b it will be x minus a square upon b minus a into c minus a. So, in that for the case b to c sorry this is for in between the reason a to b, so it will be a to b, x is more than a less than equal to b and then for if it is between b and c it will be 1 minus c minus x square upon c minus b into c minus a. So, it will be x will be more than b and less than equal to c. So, this is how the cumulative distribution function looks like. So, once you have this 3 parameters given you can find the f x value or the cumulative probability value cumulative distribution function value based on these formulas.

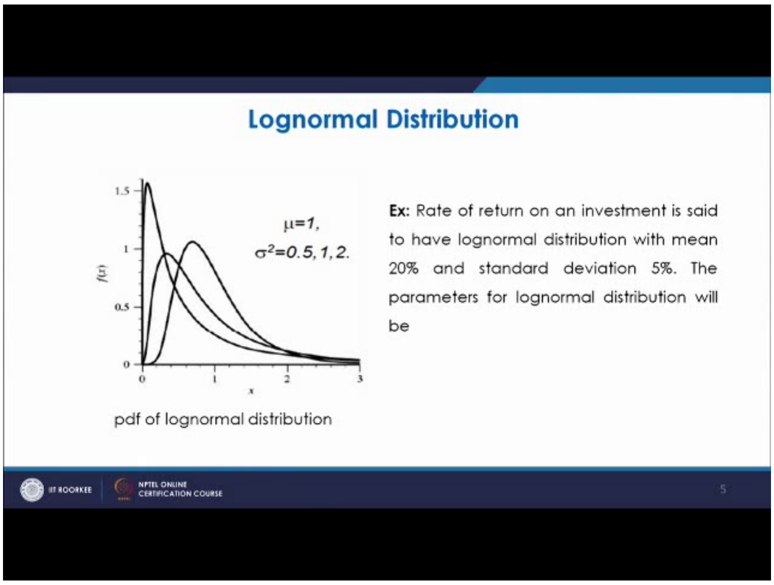
Other way you have other values also. So, what you see here is the mode is what you see is the mode is occurring at b maximum probability is value as at b. So, mode occurs at b and the mean value E x will be a plus b plus c by 3 and the mode is 3 times E x minus a plus c. So, this is nothing, but mode which occurs at b it will be 3 times expected value minus a plus c. So, this is these are the formulas for triangular distribution.

Let us see an example which talks about the time taken to execute a program which is following the triangular distribution with alpha as 0.05 millisecond, b as 1.1 millisecond and the c is 6.5 milliseconds. So, then you have to find the probability that a random program is completed in 2.5 milliseconds or less it means you have to find the F x capital F x that is cumulative distribution function value. So, that it completes between below this 2.5 milliseconds and for such examples what you find is you will find F 2.5.

So, once you find the example. So, for that F 2.5 you have to find a b and c is given once a b and c because it is occurring in the second stage between b and c the value of x is between b and c. So, F x will be, if this formula will be used a b and c value is given a is 0.05, b is 1.1 and c is 6.5. So, this value is between 1.1 and 6.5. So, we are we are going to use this value. So, that will be F 2.5 will be 1 minus 6.5 that is c minus x. So, that is 2.5 square divided by 6.5 minus 0.5 and 6.5 minus 1.1. So, that is what the values are quoted and once you compute this value this comes out to be 0.541. So, such problems which were any event is said to follow the triangular distribution and the parameters are given you can find these values I mean depending upon the situation you can find the different values.

Next comes is the lognormal distribution.

(Refer Slide Time: 21:56)



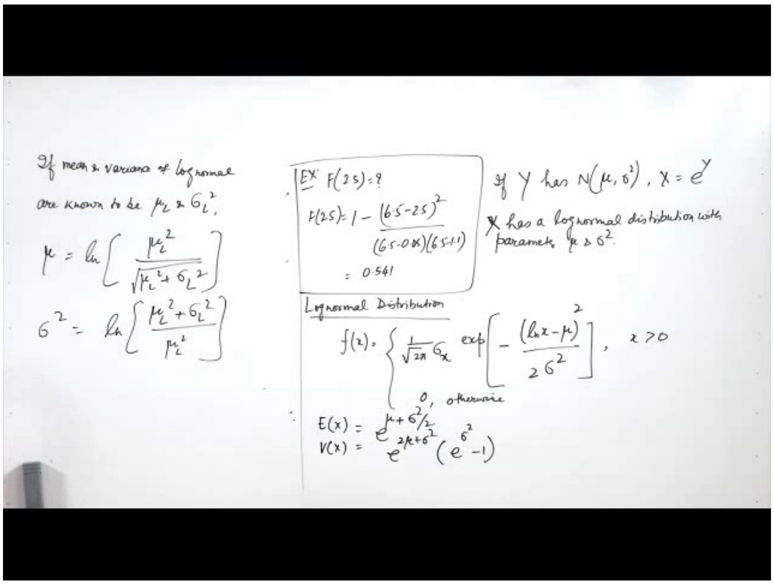
So, what is a lognormal distribution? As you see the lognormal distribution looks like this f x value is like this. So, in this case the f x if you take the lognormal distribution.

For a lognormal distribution f x is equal to 1 by root 2 pi sigma x into exponential minus of ln x minus mu square upon 2 sigma square. So, this is when x is greater than 0 and it is 0 otherwise. So, this is the probability density function for a lognormal distribution function. Now in this case there are certain parameters and there are certain you know points about this lognormal distribution you have E x comes out to be e raised to the power mu plus sigma square by 2 and the variance is coming out to be e raise to the power 2 mu plus sigma square into e raise to the power sigma square minus 1. So, this is how the mean and variance is computed for lognormal you know distribution curves. Lognormal pdfs with different values I mean, we have mean as one and different value of standard deviation or variance is shown here. So, this is we see.

Now, in this case what happens that as we see you have certain you know in this case there are certain points to be written you must know about it that this mu and sigma square it is not the mean and variance of lognormal yeah. So, if the y if any you know random variable y the there are certain traits about this lognormal variable if any random variable Y has. A normal distribution of mu and sigma square as the parameter then x equal to e y it will be having a lognormal distribution with mu and sigma square. So, it has a X has a lognormal distribution with parameter mu and sigma square. So, that is how this is the one of the property of such kind of distribution functions.

Now, if the mean and variance of lognormal are known to be mu ln sigma l square in that case mu and sigma square, so that we can further write.

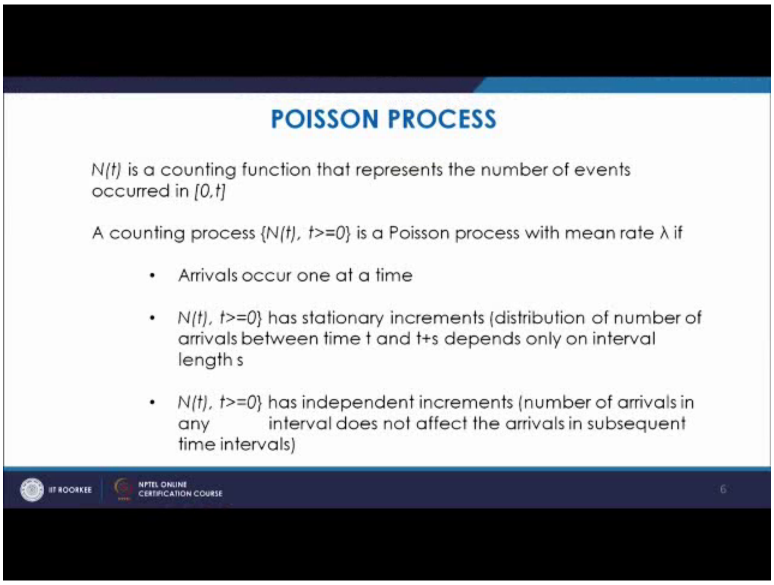
(Refer Slide Time: 26:12)



So, if mean and variance of lognormal are known to be mu L and sigma L square. So, in that case mu will be equal to log of mu l square upon under root mu L square plus sigma L square. So, this is how this mu will be calculated mu is related to that that is why it is known as lognormal and sigma square will be ln of mu L square plus sigma L square divided by mu L square. So, this is how this once you have the mean and variance of lognormal is known in that case mu and sigma square can be computed using this mu and sigma I mean sigma L for the lognormal that is mean for the lognormal and this is the variance for the lognormal and for this from that these parameters can be computed.

Next is, next one of the important process is the Poisson process.

(Refer Slide Time: 27:41)



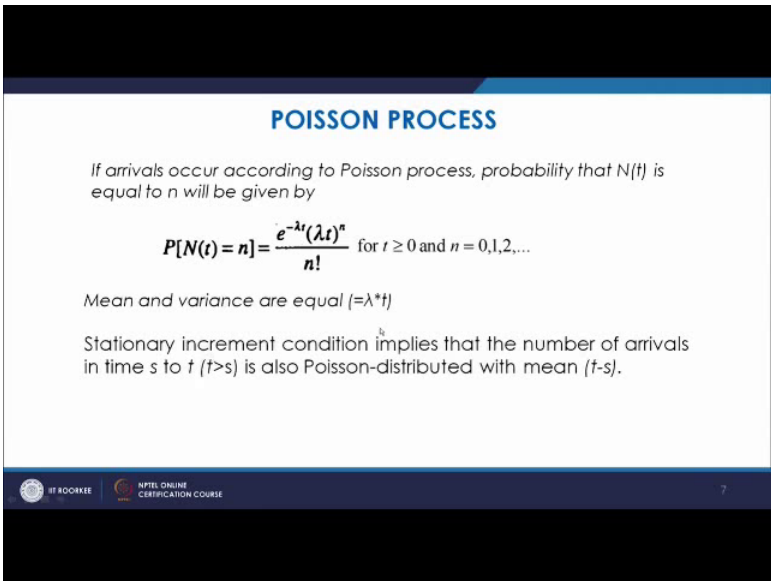
So, we have discussed about the Poisson distribution and where we try to see that the inter arrival time is normally modeled in I mean using this Poisson distribution. So, we discussed in the discrete distribution and the Poisson inter arrival time. I mean inter arrival time is modeled using that Poisson you know distribution which was a discrete distribution.

Now, what is the Poisson process? So, when we talk about the inter arrival time. So, that is basically Poisson distributed, but when we talk about the function which tells about the number of events which occur in certain time interval. So, that is known that is basically represented by such distribution. So, this in this distribution we are coming across the term N t. So, N t is basically the number of times the events have occurred in in particular time of t. So, from 0 to t how many times or between you know s to s plus t or. So, or t to t plus as means the interval is s. So, in that how many times. So, this numbers this event has occurred. So, that will talk this probability will basically we told by this Poisson process.

So, this is the trait of the Poisson process the counting process N t that is t is more than equal to 0 is a Poisson process with mean rate lambda the mean rate lambda if the arrivals occur one at a time. So, they are telling that there are these are the 3 assumptions which should be followed and then it is said to be a Poisson process. So, the condition is that arrival should occur one at a time then it has stationary increments stationary increments means a distribution of number of arrivals between time t and time t plus s depends only on interval length s. So, means the what will happen between time t and time t plus s it only depends on these interval length s does not depend upon t or so, so that is known as the it is one of the traits of this Poisson process this are the assumptions which are taken if this assumptions are satisfied then we say that it is a Poisson process.

So, and the third assumption is that for N t with where the t is more than 0, it has independent increments means number of arrivals in a in any interval does not affect the arrival in the subsequent time interval. Means that, if there are intervals and if in any interval there is more number of arrival it does not mean that it will affects the subsequent number of arrivals in the next period does not mean that in the next period there will be less number of arrivals or more number of you cannot predict about it. So, that is how it is you have these 3 you know assumptions and which are must be satisfied for being said it to be the Poisson process.

(Refer Slide Time: 31:08)



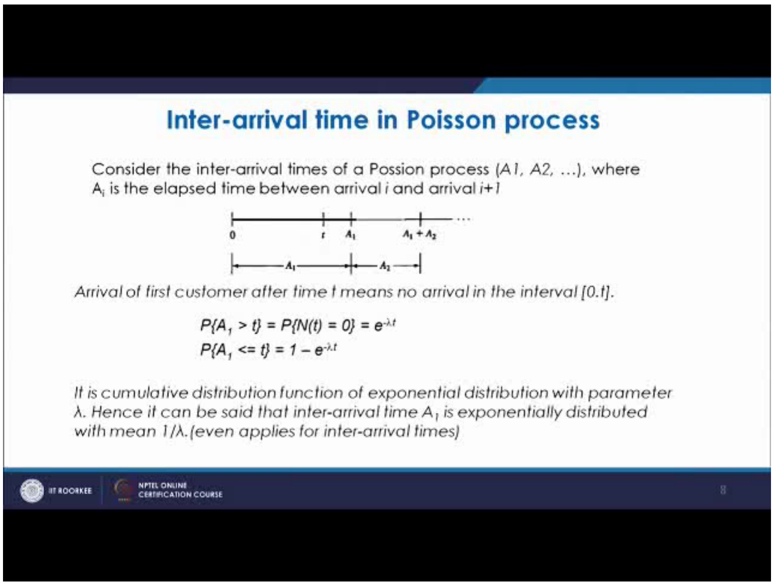
Now, if the arrivals occur according to Poisson process then the probability N t will be where N t is equal to N means if there are N arrivals for N arrivals in t time will be equal to, as we see this expression tells probability of N t is equal to N that is equal to e raised to the power minus lambda t lambda t raised to the power N divided by N factorial. So, this is the probability density function for the Poisson process.

Now, in this case mean and variance are said to be equal to lambda times t, so mean and variance both are same in this case. So, as we discussed that the stationary increment means number of arrivals in time s to t. So, if the t is t time is more than s. So, s to t will is also Poisson distributed with mean t minus s basically. So, from s to t number of arrivals in time that will also be again distributed with. So, if you do that arithmetical calculation it will basically a function of t minus s, that is how it is calculated.

So, with some of the arithmetic calculations it can be shown that while the inter arrival times the time between the 2 arrivals while that time is you know exponentially distributed or Poisson distributed the number of arrivals which occur between some time interval that is basically the Poisson process. So, while, you can further even find suppose the first arrival has occurred after t time means there is no arrival before t. So, from there if you do certain calculations you can find that the time to first arrival is basically that is the inter arrival time between the first arrival if you find that that will be basically the Poisson distributed. So, that will be exponential you know function.

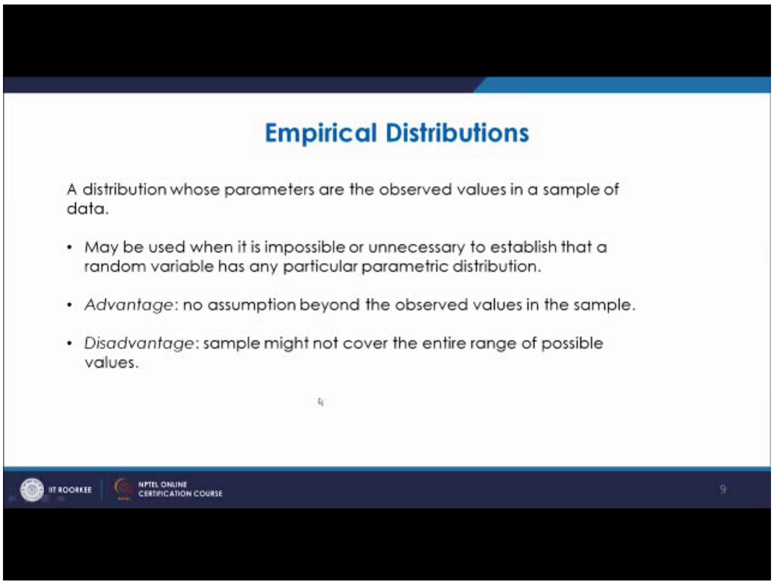
So, basically from there you can find that these interarrival times are while the Poisson distributed this is the Poisson. So, exponentially distributed this is known as this the total interval and number of arrivals come I mean in a particular time interval is Poisson process, that can be found out by using this probability distribution function.

(Refer Slide Time: 33:54)



So, as we here that if suppose we are talking about inter arrival times of a Poisson process and if there is a, there is one in arrival which is occurring after time t means arrival if the A 1 is more than t, in this case now in that case first. So, time t means there is no arrival before time t and for that the probability if you try to find it will be with 0 that arrival it will be one minus e raised to the power minus lambda t. So, it is nothing, but the cumulative distribution function of the exponential distribution with parameter lambda. And that is why it see said that the inter arrival time is exponentially distributed and then this is Poisson distributed I mean the total number of arrivals in between certain time interval.

(Refer Slide Time: 34:55)



So, it has basically a lot of use when we talk about the queuing problems and there you find other parameters.

Then comes the empirical distribution functions. As we know the empirical distribution function means here you have the observed values every time you not necessarily you will find the data I mean it which is following certain standard distribution function, it may follow any particular distribution any distribution function which is not of this standard type. So, in those cases what you have to do is either be discrete or continuous in the case of continuous or in the case of discrete for discrete values you have to find the relative frequency you have to find a cumulative frequency based on the relative frequency you can find the pdf or pmf graph or based on the cumulative distribution you know cumulative relative frequency, you can find the cumulative distribution curve.

Similarly, in the case of you know continuous distribution functions you have the ranges for that you will have certain relative probabilities and based on that you can find the cumulative relative probabilities, cumulative relative frequencies and based on that for that particular ranges you can have the graph and a cdf cumulative distribution function can be drawn. So, that is known as empirical distributions which does not follow any standard type of distribution function. We can discuss about these by solving a problem in the coming lecture and let us talk about some of the traits of this it may be used when it is impossible to or unnecessary to establish a random variable having any particular parametric distribution.

So, if it does not follow we should not be bothering much about it then advantage is that no assumption beyond the observed values and the disadvantage is that it may not cover the entire range of possible values. So, this is how the empirical distribution functions are defined and we can discuss we can see by example in the solve example problems.

Thank you.